




# Insights into the transcriptional and translational mechanisms of linear organellar chromosomes in the box jellyfish *Alatina alata* (Cnidaria: Medusozoa: Cubozoa)

Ehsan Kayal, Bastian Bentlage & Allen G. Collins


To cite this article: Ehsan Kayal, Bastian Bentlage & Allen G. Collins (2016) Insights into the transcriptional and translational mechanisms of linear organellar chromosomes in the box jellyfish *Alatina alata* (Cnidaria: Medusozoa: Cubozoa), RNA Biology, 13:9, 799-809, DOI: 10.1080/15476286.2016.1194161


To link to this article: <https://doi.org/10.1080/15476286.2016.1194161>

 View supplementary material 


 Published online: 28 Jul 2016.

 Submit your article to this journal 

 Article views: 985

 View related articles 

 View Crossmark data 

 Citing articles: 1 View citing articles 

RESEARCH PAPER

## Insights into the transcriptional and translational mechanisms of linear organellar chromosomes in the box jellyfish *Alatina alata* (Cnidaria: Medusozoa: Cubozoa)

Ehsan Kayal<sup>a</sup>, Bastian Bentlage<sup>a</sup>, and Allen G. Collins<sup>a,b</sup>

<sup>a</sup>Department of Invertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington DC, USA; <sup>b</sup>National Systematics Laboratory of NOAA's Fisheries Service, National Museum of Natural History, Washington, DC, USA

### ABSTRACT

Background: In most animals, the mitochondrial genome is characterized by its small size, organization into a single circular molecule, and a relative conservation of the number of encoded genes. In box jellyfish (Cubozoa, Cnidaria), the mitochondrial genome is organized into 8 linear mito-chromosomes harboring between one and 4 genes each, including 2 extra protein-coding genes: *mt-polB* and *orf314*. Such an organization challenges the traditional view of mitochondrial DNA (mtDNA) expression in animals. In this study, we investigate the pattern of mitochondrial gene expression in the box jellyfish *Alatina alata*, as well as several key nuclear-encoded molecular pathways involved in the processing of mitochondrial gene transcription. Results: Read coverage of DNA-seq data is relatively uniform for all 8 mito-chromosomes, suggesting that each mito-chromosome is present in equimolar proportion in the mitochondrion. Comparison of DNA and RNA-seq based assemblies indicates that mito-chromosomes are transcribed into individual transcripts in which the beginning and ending are highly conserved. Expression levels for *mt-polB* and *orf314* are similar to those of other mitochondrial-encoded genes, which provides further evidence for them having functional roles in the mitochondrion. Survey of the transcriptome suggests recognition of the mitochondrial tRNA-Met by the cytoplasmic aminoacyl-tRNA synthetase counterpart and C-to-U editing of the cytoplasmic tRNA-Trp after import into the mitochondrion. Moreover, several mitochondrial ribosomal proteins appear to be lost. Conclusions: This study represents the first survey of mitochondrial gene expression of the linear multi-chromosomal mtDNA in box jellyfish (Cubozoa). Future exploration of small RNAs and the proteome of the mitochondrion will test the hypotheses presented herein.

### ARTICLE HISTORY

Received 1 April 2016  
Revised 18 May 2016  
Accepted 23 May 2016

### KEYWORDS

*Alatina alata*; linear mitochondrial DNA; mitochondrial gene expression; mitochondrial polymerase; tRNA C-to-U editing

### Background

The mitochondrion, largely responsible for cellular respiration, is thought to have originated through a single endosymbiotic event of an  $\alpha$ -proteobacterium<sup>1</sup> very early in the evolution of eukaryotes (though see Pittis and Gabaldón (2016)<sup>2</sup> for an alternative hypothesis). Subsequent gene transfer to the nucleus of the host cell and loss of genes from the proto-mitochondrion<sup>3</sup> has rendered the typical mitochondrial genome a small molecule that encodes only a fraction of the genes necessary for the functioning and maintenance of the mitochondrion. In most animals (Metazoa), mitochondrial DNA (mtDNA) comprises a single, circular and compact molecule of ~17 kbp in size, encoding 13 (+2/–2) protein-coding genes involved in oxidative phosphorylation, 2 rRNA (rRNA) genes, and up to 25 tRNA (tRNA) genes involved in the translation of mitochondrial-encoded proteins. While gene order in the mt genome is highly variable, most protein-coding and ribosomal genes are interspersed with tRNA genes. The replication of the typical circular mtDNA is performed through the formation of a D-loop, where the replication of the two strands is out of sync.<sup>4</sup> The “tRNA punctuation model” proposes that mtDNA is transcribed into one or two polycistronic precursor transcripts (pre-mRNA molecules) that are then excised at the

tRNA sites,<sup>5</sup> releasing mostly single-gene (monocistronic) mRNAs, rRNA and tRNA transcripts.<sup>6</sup>

In the early branching metazoan phylum Cnidaria (sea anemones, corals, hydras, hydroids, and jellyfish), the mt genome encodes for only one or 2 tRNA genes,<sup>7–17</sup> challenging the tRNA punctuation model of gene expression. Species from the cnidarian class Anthozoa (the clade containing reef-building corals, sea anemones, soft corals, and their relatives) possess relatively large, circular mitochondrial genomes in which protein-coding genes are interspaced with intergenic regions (IGRs) that might play a role in mitochondrial mRNA maturation (e.g.<sup>18,19</sup>). The mt genome in the cnidarian subphylum Medusozoa (e.g. hydroids and jellyfish) is organized into mono-, bi-, or multi-linear chromosomes.<sup>7–13,20</sup> Medusozoan mtDNA is very compact with a few small IGRs.<sup>12,15</sup> In hydrozoans, Kayal *et al.* (2015)<sup>15</sup> suggested that both rRNA genes and stemloop-like secondary structures formed by IGRs could be recognized by the nucleolytic processing mechanism, producing mature often multi-cistronic mRNA transcripts.

Cubozoans (box jellyfish) have mtDNA made up of eight small linear chromosomes.<sup>12,13,20</sup> Mitochondrial chromosomes range from 4 to 6 kbp in length and harbor between one and four genes, flanked by long and conserved telomere-like inverted repeats (IRs) with opposite orientation thought to be

involved in the maintenance of the linear mito-chromosomes.<sup>13</sup> In addition to the 13 protein-coding genes common in Metazoa, most medusozoan mtDNA (excluding hydroidolinan hydrozoans) encodes for two extra protein-coding genes (*mt-polB* and *orf314*). Mohammadou et al (2004)<sup>21</sup> have found that *mt-polB* has affinities with polymerases from the B family. Kayal et al. (2012)<sup>12</sup> have suggested that *orf314* might be acting as a terminal protein by binding the 5' single-stranded end of the linear chromosomes as described for the liverwort *Marchantia polymorpha*<sup>22</sup> and some yeast species.<sup>23,24</sup> *mt-polB* and *orf314* are present in all medusozoan subclasses and lack internal stop codons; in cubozoans *mt-polB* and *orf314* are found together on an independent linear mito-chromosome. The maintenance of these genes through long evolutionary times (acquired before the divergence of Medusozoa and maintained in several medusozoan clades) has been used as evidence to suggest that they have functional roles in the mitochondrion.<sup>12,13</sup> Additional evidence on the expression of these genes is needed to confirm and refine this hypothesis and to assess specific functions.

Here, we surveyed recently released RNA-seq data from the box jellyfish *Alatina alata* (SRR1952741) and assembled the near complete multi-chromosomal mitochondrial genome, as well as several mitochondrial-targeted genes involved in mtDNA expression, including but not limited to mitochondrial aminoacyl-tRNA synthetases (mt-aaRS), polymerases, and ribosomal proteins. Our analysis offers insights into the expression of mtDNA genes and provides additional evidence about the functional roles of *mt-polB* and *orf314*.

## Material and methods

We obtained a recently published transcriptome assembly for *Alatina alata*<sup>25</sup> as well as 2 publicly available short read archive (SRA) datasets from the National Center for Biotechnology Information (NCBI; [ncbi.nlm.nih.gov/sra/](http://ncbi.nlm.nih.gov/sra/)): SRR217435 containing SOLID DNA-seq reads and SRR1952741 containing ILLUMINA HiSeq 2500 RNA-seq reads. The RNA-seq library for the data set used in this study was prepared using the TruSeq mRNA enrichment protocol, where oligo-dT beads capture RNA molecules with poly(A) tail, before random reverse transcription is performed.<sup>25</sup> Tissue on which the transcriptome assembly was based did not include any gut tissue, lessening concerns about contamination from other organisms. We mapped the SOLID DNA-seq data to a published mt genome of *Alatina alata* (JN642330-JN642344) using Bowtie2<sup>26</sup> with *-end-to-end* and *-very-sensitive* options to estimate individual chromosome and gene coverage.

Fig. S1 details the bioinformatics workflow we used in this study. In short, we captured and assembled mitochondrial contigs from the assembled transcriptome using the BLAST plugin in Geneious v.9.<sup>27</sup> We filtered raw reads from the ILLUMINA HiSeq RNA-seq data with Trimmomatic v.0.32<sup>28</sup> and mapped filtered reads to the assembled mitochondrial chromosomes using Bowtie2, with default settings. We used Geneious to visualize and manually check the assemblies, and to build consensus sequences. We annotated individual chromosomes following the reference mitogenome and predicted the origin and end of transcription. Finally, we mapped trimmed RNA-

seq reads back to the mitochondrial consensus sequences to estimate individual chromosome and gene coverage with Bowtie2 using *-end-to-end* and *-very-sensitive* options (Fig. S1).

We also investigated several genes involved in the expression and function of the mitochondrion from the RNA-seq data: the glutamyl-tRNA amidotransferase subunits (GatA-C) that are absent in other cnidarian classes,<sup>29</sup> the mitochondrial ribonucleases complex P (MRPP1-3) and Z (RNase Z) responsible for the cleavage of the 5' and 3' ends of the tRNA, the mitochondrial poly(A) RNA polymerase (MTPAP) that creates the 3' poly(A) tails of mitochondrial transcripts, the full set of aminoacyl-tRNA synthetases (aaRSs), the mitochondrial DNA-directed RNA polymerase (POLRMT), and mitochondrial ribosomal proteins involved in transcription. First, we identified homologs of the genes of interest in several species from the Uniprot database (<http://www.uniprot.org/>; last accessed 3/5/2016; the list of query sequences is found in Table S1). We then used the BLAST plugin in Geneious to search the transcriptome assembly for *A. alata* for putative homologs. All BLAST hits with a max e-value of 1e-1 were retained. Retained hits were aligned and assembled when possible into consensus contigs. The resulting contigs were individually checked against the NCBI protein database (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) using blastx to remove potential bacterial contaminants (the top 10 or more hits of bacterial origin) and genes not involved in mitochondrial expression (eukaryote hits with the majority reporting other functions). When the identity of the gene was dubious (not a majority reporting a single function), the translated coding region was blasted against the Uniprot database. Incomplete transcripts were extended, when possible, to include parts of the 5' and 3' untranslated regions (UTRs) using a combination of the mirabait script from MIRA v.4,<sup>30</sup> MITObim v.1.8,<sup>31</sup> and the overlap-layout-consensus assembler implemented in Geneious. The presence of putative mitochondrial targeting-signal peptides was assessed using the online versions of the programs MitoProt II,<sup>32</sup> TargetP v1.1,<sup>33</sup> Predotar<sup>34</sup> and Mitofates.<sup>35</sup> Given that the bioinformatics tools for signal peptide prediction are trained on only a handful of model organisms, we considered a signal peptide present when probabilities were >0.5 for at least 2 of the 4 prediction programs. All identified mitochondrion-targeted genes were deposited in the FigShare digital repository and can be accessed at <https://dx.doi.org/10.6084/m9.figshare.3380848>.

Since two out of 3 components of the mitochondrial ribonucleases complex P (MRPP1-2) are part of large gene families, we used a phylogenetic approach to confirm the identifications of all 3 mitochondrial ribonucleases complexes (MRPP1-3). This approach also allowed us to test the accuracy of our BLAST-based method described above. We first downloaded the sequences that matched “trna methyltransferase 10” and “reviewed:yes” for MRPP1 (240 sequences), “3 hydroxyacyl coa dehydrogenase” and “reviewed:yes” for MRPP2 (282 sequences), and “mitochondrial ribonuclease protein p” for MRPP3 (136 sequences) from the Uniprot database (last accessed 4/5/2016). We then aligned each set with the corresponding sequence identified from *Alatina alata* using the online program MAFFT v.7<sup>36</sup> with the L-INSi algorithm. Gblocks v0.91b<sup>37</sup> was used on *mrrp2* with gap allowed at all positions; the *MRRP1* and *MRRP3* alignments were not filtered since

Gblocks would remove most (MRPP1) or all (MRPP3) the data given the high level of sequence variability within the alignments. Phylogenetic analyses were performed under a Maximum Likelihood framework using RAXML v.8.2.7<sup>38</sup> with the PROTGAMMAWAG model of sequence evolution (100 independent searches for the best tree; 1000 bootstraps). All alignments and trees are deposited in the FigShare digital repository and can be accessed at <https://dx.doi.org/10.6084/m9.figshare.3380848>.

We obtained the full set of sequences belonging to the polymerase  $\alpha$  superfamily (*DNA\_pol\_A* PF00476) from the Pfam database<sup>39</sup> and created a local HMM database using HMMER v3.1b1.<sup>40</sup> We searched the transcriptome assembly using HMMER and isolated contigs matching the gamma polymerase gene (*polG*). We used Geneious to assemble the positive matches into a single contig and checked the assembly by mapping the filtered reads against the *polG* contig with Bowtie2 (see above).

## Results and discussion

### Expression of mt genes in *Alatina alata*

We mapped SOLID DNA-seq data from the SRR archive (SRR217435) to published mitochondrial chromosomes obtained from these data (JN642330-JN642344<sup>13</sup>) and found that all eight mito-chromosomes have similar but unequal (variance 18%) read coverage. That is, no chromosome is found to have considerably more coverage than the others (Fig. 1A; Table S2). Interestingly, we found the genes *cox1*, *cox2*, *mt-polB*, *nad4* and *orf314* having the highest coverage in the DNA-seq SOLID data (Table 1). The variation in the coverage between various mito-chromosomes in the DNA-seq SOLID data does not appear to correspond to the functional requirements of mitochondrial genes. For instance, despite the key role played by mt-rRNAs, *rnl* and *rns* display no noticeable deviation from the average read coverage, in stark contrast with RNA-seq coverage (Fig. 1A, Table 1). It is very likely that all eight mito-chromosomes are in equimolar proportion in the mitochondrion of *A. alata*.

We recovered nearly complete mitochondrial chromosomes from the RNA-seq reads, including part of the terminal inverted repeats (IRs). Compared to the reference mitogenomes, 10 bp are missing in the RNA-based assemblies, which we define as the origin of transcription (OT). The OT is part of a 30 bp region present on all mito-chromosomes with putative stem-loop structures predicted by the DNA folding program from the mfold Web Server with default parameters<sup>41</sup> (Fig. 2). The complete conservation of the OT and its putative secondary structure support its role as the origin of transcription for the mito-chromosomes. Smith et al (2012)<sup>13</sup> described a gene-conversion model for the presence of partial duplication at the subtelomeric regions of all mito-chromosomes. The level of conservation of subtelomeric regions varies among mito-chromosomes: the coding regions of seven out of the eight mito-chromosomes start with the 5' end of *rnl*, with the highest conservation found in *nad2* (Fig. S2). A small (~5 bp) sequence conservation was found between the 5' end of *rnl* and *rns*.<sup>13</sup>

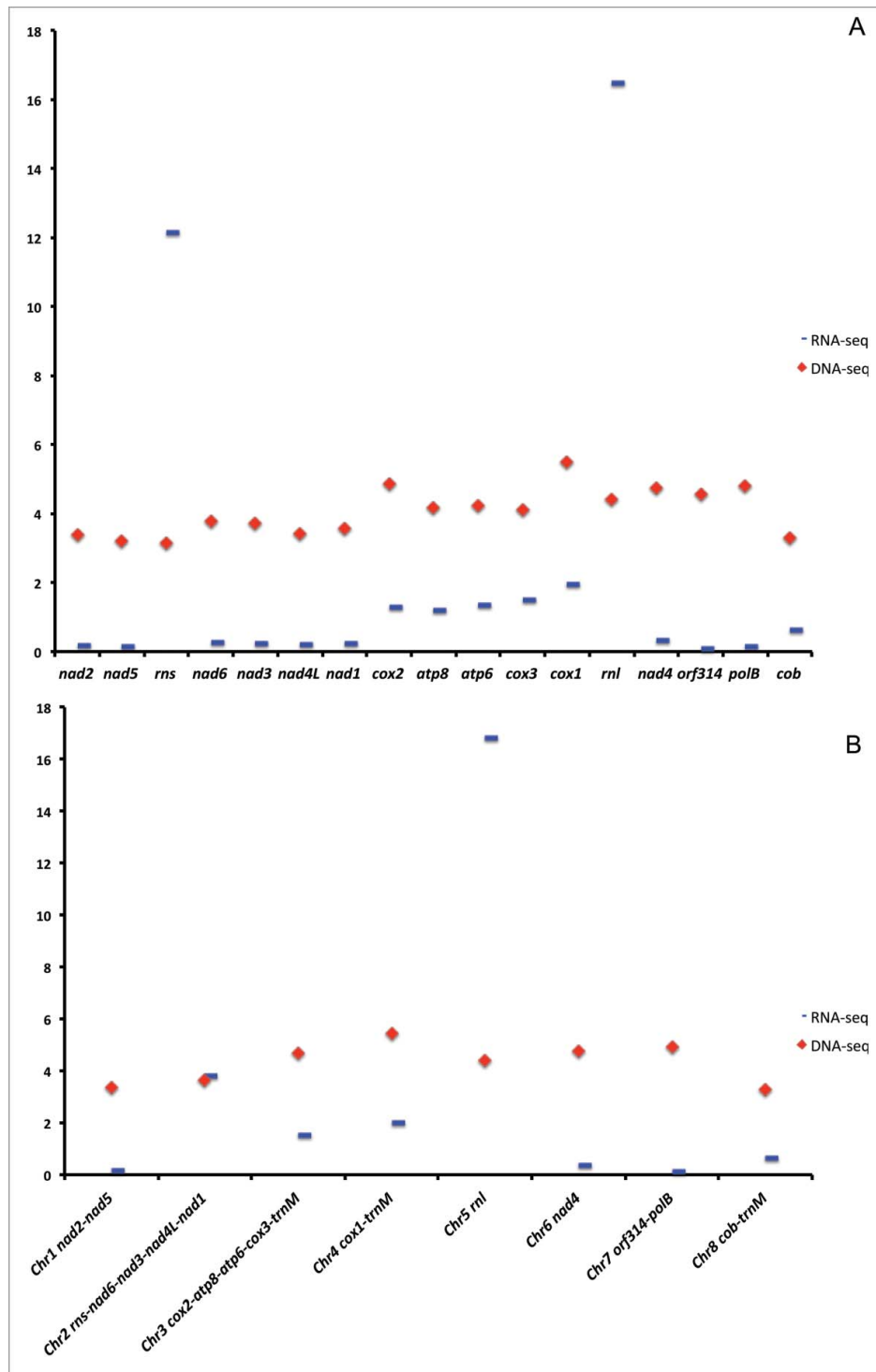
The pattern of gene expression corresponds to the organization of the genes on each chromosome, with no break within mitochondrial contigs, as described in other medusozoans,<sup>15</sup> where  $\geq 10$  bp stem-loop structures might be responsible in mitochondrial pre-mRNA processing. Such a pattern suggests that mito-chromosomes are expressed into single per-chromosome transcripts (Fig. 3A) starting at the OT region a few base pairs before the IR upstream of the coding region, and ending 161 bp within the IR downstream of the coding region. Average coverage ranged from ~130 to ~16,800 nucleotides per base of the unique regions of the genes (the large regions of genes not duplicated on other chromosomes), and the highest values were observed for the mito-chromosomes harboring the rRNA genes (Fig. 1B; Table 1). High-level representation of mt-rRNA in the RNA-seq data is expected given that ribosomes are essential in the expression of the mitochondrial genes. Interestingly, the 2 chromosomes containing *cox1-3* and *atp6* and 8 genes are also highly expressed in the RNA-seq data (Fig. 1B; Table S2).

Conservation of subtelomeric regions suggests that transcription can start on any mito-chromosome and an as yet unknown elongation factor is required to recognize the correct gene (Fig. 3A). Mitochondrial ribosomal proteins or rRNA genes might act as elongation factors in the transcription. Mito-chromosome 2 harbors *rns* along with four protein-coding genes (*nad1*, 3, 4L and 6), and displays the highest coverage of RNA reads (mean coverage ~4 kbp; Fig. 1B). Read coverage is not evenly distributed along this mito-chromosome: the protein-coding genes have read coverage ~55 times lower than *rns* (Table 1). The dramatic drop in read coverage is observed at the 3' end of *rns*, which can fold into a stem-loop structure (Fig. S2). We propose that the secondary structure at the 3' end of *rns* is recognized by some factor(s) and helps regulate transcription by detaching the RNA POLRMT from its template.

We found no read spanning beyond one-fourth (161 out of 702 bp) of the IR downstream of the coding region of the chromosomes. Studies have shown polyadenylation of both rRNA and mRNA in mammalian mitochondria.<sup>42</sup> In the transcriptome of *Alatina alata*, we identified and annotated several complete sequences (including both the 5' and 3' UTR regions) homologous to the poly(A) polymerase (PAP) gene, one of which (MTPAP PIP5K1A) likely contains an N-terminal mitochondrial-targeting signal peptide (Table 2). Assuming similar polyadenylation is taking place in the mitochondria of all animal species, we posit that the recognition site for MTATP in *Alatina* is about 1/4<sup>th</sup> of the way into the IR of the mitochondrial chromosomes.

### A functional mitochondrial-encoded type B DNA polymerase

We recovered the complete sequence of the *polG* gene whose product is responsible for mtDNA replication, including the mitochondrial-targeting signal peptide (Table 2). The product of this gene is similar in size to its counterparts in other animals. Medusozoan mitogenomes (excluding hydrozoans) harbor two additional protein-coding genes *orf314* and *mt-polB*. Our previous studies<sup>12,13</sup> have suggested that *orf314* and *mt-polB* play roles in the maintenance and replication of the mitochondrial telomeres: the product of *mt-polB* has affinities with members of the DNA-directed DNA



**Figure 1.** Mitochondrial read coverage values in the mtDNA of *Alatina alata*. (A) Nucleotide coverage of the unique regions of individual mitochondrial genes from the DNA-seq and RNA-seq data. (B) Nucleotide coverage of the unique regions of mitochondrial chromosomes from the DNA-seq and RNA-seq data. The Y axis represents mean coverage in kbp.

polymerase type B found in plant and fungal mitochondria, and viruses; *orf314* was hypothesized to encode for a DNA-binding protein that could sit at the 5' single-stranded end of mitochondrial telomeres to prevent their degradation by exonucleases and may be involved in the maintenance of the integrity of the mito-chromosomes during replication.<sup>7,12</sup> It has been posited that the linearization of medusozoan mtDNA originated in association with the incorporation of a

linear plasmid containing both *mt-polB* and *orf314*.<sup>12</sup> The presence of these genes on a separate mito-chromosome in *A. alata* was already suggested as evidence of positive selection for their maintenance.<sup>13</sup> We recovered the chromosome containing these genes from the RNA-seq data with per-gene coverage values similar but below that of the other protein coding genes (Table 1 and S2), indicating that these genes are expressed and polyadenylated (i.e., present in the mRNA



**Table 1.** Read coverage of the uniq regions of the mitochondrial genes in the DNA-seq and RNA-seq datasets.

mt gene	RNA-seq to shortened gene					DNA-seq to shortened gene				
	size (bp)	read* (x10 <sup>3</sup> )	nucl # (x10 <sup>3</sup> )	read cov <sup>§</sup>	nucl cov <sup>§</sup>	size (bp)	read (x10 <sup>3</sup> )	nucl (x10 <sup>3</sup> )	read cov	nucl cov
<i>nad2</i>	1199	2.0	196.7	2	164	1198	81.5	4076.2	68	3403
<i>nad5</i>	1591	2.4	237.2	1	149	1591	102.3	5115.2	64	3215
<i>rns</i>	896	108.8	10890.6	121	12155	896	56.4	2819.6	63	3147
<i>nad6</i>	537	1.3	132.8	2	247	537	40.6	2030.9	76	3782
<i>nad3</i>	351	0.8	79.1	2	225	351	26.0	1301.0	74	3707
<i>nad4L</i>	291	0.6	55.8	2	192	291	19.9	995.2	68	3420
<i>nad1</i>	967	2.1	215.0	2	222	969	69.1	3458.2	71	3569
<i>cox2</i>	719	9.2	920.8	13	1281	698	68.0	3402.5	97	4875
<i>atp8</i>	210	2.5	253.3	12	1206	210	17.5	874.9	83	4166
<i>atp6</i>	705	9.5	950.3	13	1348	705	59.5	2977.2	84	4223
<i>cox3</i>	679	10.2	1019.2	15	1501	679	55.9	2797.3	82	4120
<i>cox1</i>	1535	30.0	3001.6	20	1955	1535	168.3	8417.8	110	5484
<i>rnl</i>	1471	243.4	24238.7	165	16478	1473	129.8	6489.1	88	4405
<i>nad4</i>	1358	4.4	442.6	3	326	1360	129.0	6453.4	95	4745
<i>orf314</i>	294	0.2	24.6	1	84	294	26.9	1342.7	91	4567
<i>polB</i>	957	1.3	126.0	1	132	1002	96.3	4815.2	96	4806
<i>cob</i>	1003	6.3	628.9	6	627	1003	66.1	3305.0	66	3295

read\*: raw count of read; nucl #: raw count of nucleotide; read cov<sup>§</sup>: read coverage; nucl cov<sup>§</sup>: nucleotide coverage.

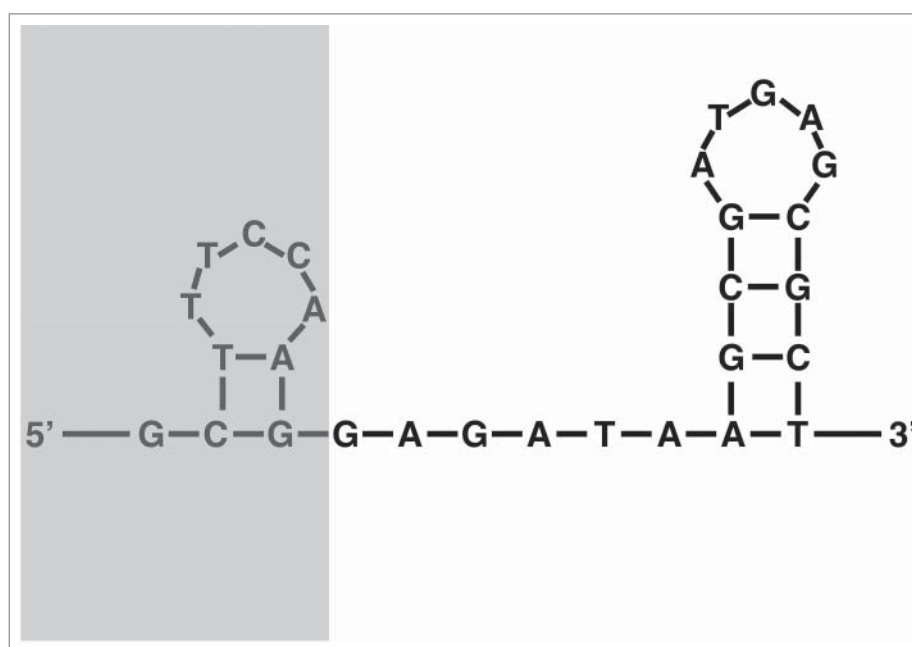
pool). These data provide further evidence that both *orf314* and *mt-polB* are functional in the mitochondrion.

### Near-parallel loss of mt-tRNA and mt-aaRS in Cubozoa suggests tRNA editing

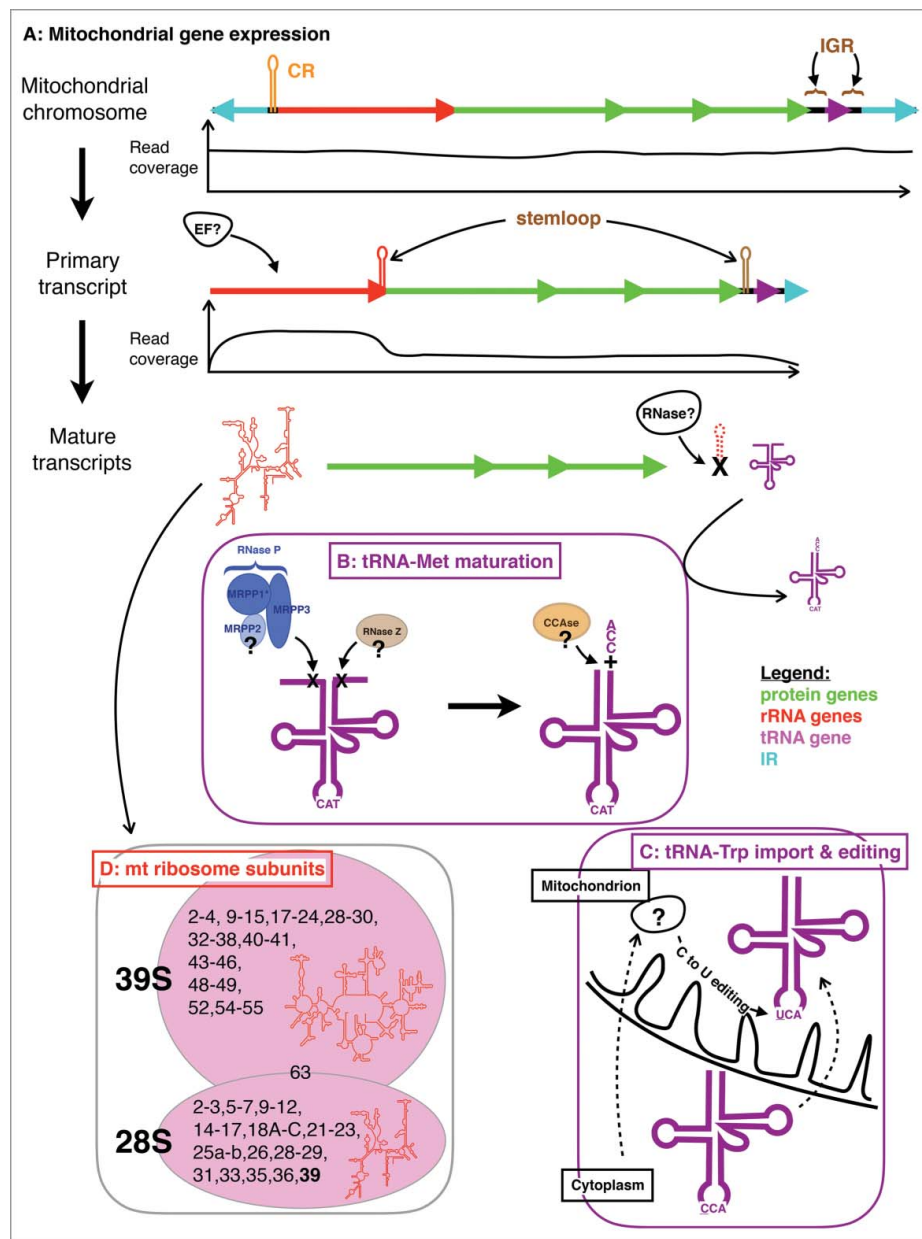
The only complete mitochondrial genome for Cubozoa published to date, that of *Alatina alata* (JN642330-JN642344), encodes one single mitochondrial tRNA for methionine (mt-tRNA<sup>Met</sup> or *trnM*),<sup>13</sup> suggesting that the other tRNAs are imported from the cytoplasm. We recovered the complete set of cytoplasmic aminoacyl-tRNA synthetases (nc-aaRS) but only one mitochondrial aminoacyl-tRNA synthetase (mt-PheRS) from the RNA-seq data (Table 3). This pattern is in accord with the evolutionary scenario of parallel loss of mt-

tRNAs and mt-aaRS, excluding mt-PheRS, which is made of a single protein as opposed to nc-PheRS composed of 2 subunits encoded by 2 separate genes.<sup>29,43</sup> Three full transcripts (Glu/ProRS, IleRS and ValRS) did not harbor a recognizable 5' mitochondrial-targeting signal peptide (Table 3), while the sequence for HisRS was missing the 5' end where the mt-targeting signal peptide is expected to be found. The absence of mt-targeting signal peptides in Glu/ProRS, IleRS and ValRS suggests some post-transcriptional modification that might be responsible for the transport of these nc-aaRSs into the mitochondrion, though the possibility of a deviant targeting signal not identified by the prediction programs cannot be ruled out.

Previous studies have suggested that the loss of mt-tRNAs preceded that of the corresponding mt-aaRSs in Cnidaria<sup>43</sup>; cases where an mt-tRNA is present in the absence of the corresponding



**Figure 2.** Secondary structure of the putative origin of translation (OT) found on all 8 chromosomes of the mt genome of *Alatina alata* using the DNA folding program from the mfold Web Server with default parameters. The greyed area is present at the DNA level but absent from the RNA-seq data.



**Figure 3.** Gene expression in the mitochondrion of *Alatina alata*. (A) Pattern of gene expression and their coverage in the DNA-seq and RNA-seq data sets for a model mito-chromosome. Graphs of read coverage do not correspond to the true values, but rather approximate graphical representations for visualization purposes. CR = putative control region; IGR = inter-genic region; EF? = putative elongation factor involved in the expansion of the rRNA; RNase? = putative enzyme responsible for recognizing and excising stemloop structures found at the 3' end of rRNA and 5' of tRNA genes. Colored arrows correspond to protein coding genes (green), ribosomal gene (red), tRNA gene (purple), and inverted repeats or IR (cyan). (B) Nuclear genes involved in the maturation of the mitochondrial tRNA-Met: RNase P complex made of three subunits MRPP1-3 responsible for excising extra nucleotides at the 5' end of the tRNA; RNase Z responsible for excising extra nucleotides at the 3' end of the tRNA; CCAs responsible for adding CCA to the 3' end of the tRNA. (C) Import and editing of cytoplasmic tRNA-Trp, where the CCA anticodon is edited via an as yet unknown C-to-U editing enzyme (?). (D) Composition of the mitochondrial ribosome large (39S) and small (28S) subunits. Secondary structures of the large (*ml*) and small (*ms*) rRNA subunits are in red; numbers correspond to individual mitochondrial ribosomal proteins (MRPs) with identified (in black) or lacking (in red) mitochondrial-targeting signal peptides. Number in bold is the newly identified mitochondrial ribosome small subunit S39.

mt-aARS can be explained by recruitment of the mt-tRNA for other functions, such as a structural component of the mitochondrial ribosome.<sup>29</sup> In mammals, where mt-tRNA<sup>Val</sup> has been found as a structural part of the mitochondrial ribosome, the tRNA gene (*trnV*) is encoded between the small and large ribosomal subunits (*rns* and *rnl* respectively) in the mt genome, leading to a polycistronic transcript containing all three RNAs.<sup>44,45</sup> In demosponge mtDNA, *trnG* (mt-tRNA<sup>Gly</sup>) and *trnV* are the most common tRNA genes found between *rns* and *rnl*, but never *trnR* (mt-tRNA<sup>Arg</sup>),<sup>46</sup> yet mt-ArgRS is one of the five mt-aARSs (the others

being mt-CysRS, mt-LysRS, mt-ThrRS and mt-ValRS) that seem to have been lost in this clade.<sup>29</sup> This pattern cannot explain the role of these tRNAs as structural parts of the ribosome as shown in mammals.<sup>44</sup> The presence of mt-tRNA<sup>Met</sup> on three out of the eight mitochondriomes of *A. alata* concurrent with the absence of mt-MetRS in the nucleus is an interesting pattern that cannot be explained by the role of this tRNA as a structural component of the mitochondrial ribosome.

We investigated several key genes responsible for mitochondrial tRNA processing, namely RNase P (for 5'-end

**Table 2.** Read coverage of some mitochondrial-targeted genes in the RNA-seq data set.

gene	CDS size <sup>¶</sup>	mt-signal	mt-sign size <sup>§§</sup>	MitoProtII	TargetP	Predotar	MitoFates
<i>polG</i>	3429	+	20/35	<b>0.99</b>	<b>0.88</b>	<b>0.93</b>	<b>0.79</b>
<i>CCAsel</i>	1494	–	33	<b>1.00</b>	0.37	0.02	0.30
<i>ELAC2a</i>	>1134	?	?	?	?	?	?
<i>ELAC2b</i>	>1612	–	39	<b>0.9945</b>	0.443	0.11	0.126
<i>MRRP1</i>	1134	+	15/–	<b>0.86</b>	0.39	<b>0.51</b>	<b>0.71</b>
<i>MRRP2</i>	774	–	14/69	0.41	<b>0.65</b>	0.13	0.27
<i>MRRP3</i>	1941	+	–/48	<b>0.97</b>	<b>0.88</b>	<b>0.75</b>	0.04
<i>MTPAP PIP5K1A</i>	2160	+	16/9	<b>0.90</b>	<b>0.68</b>	<b>0.63</b>	0.20
<i>POLRMT</i>	3681	+	47/18	<b>0.98</b>	<b>0.79</b>	0.07	0.28
<i>PTCD1</i>	1527	–	–/–	0.1953	0.183	0.07	0.464

Presence of mt-signal is assumed when at least two prediction software display probabilities >0.50 (in bold). CDS size<sup>¶</sup>: size of the coding regions found in the transcriptome assembly; mt-sign size<sup>§§</sup>: number of amino acids composing the mt-targeting signal peptide as predicted by MitoProt II v1.01 and TargetP v1.1. Probability of mitochondrial targeting were obtained using MitoProt II v1.01, TargetP v1.1, Predotar v.1.03 and MitoFates.

maturation), RNase Z (for 3'-end maturation), and CCAsel (responsible for adding CCA to the 3'-end of the tRNA) (Table 2). In humans, RNase P is a protein complex composed of three subunits: MRPP1, one out of at least three homologs of a tRNA methyltransferase 10 (C homolog); MRPP2, a 3-hydroxyacyl-CoA dehydrogenase; MRPP3, the mitochondrial ribonuclease P protein 3.<sup>47,48</sup> We identified homologs to both MRPP2 and MRPP3 (the latter with a mitochondrial targeting signal peptide) but not to MRPP1 (Table 2). Instead, our phylogenetic analysis identified a tRNA methyltransferase 10 A homolog (Fig. S3) with a mitochondrial targeting signal peptide, while human homologs do not harbor such a signal peptide (data not shown). We therefore decided to label the

transcript tRNA methyltransferase 10 homolog A as MRPP1 (Table 2). Phylogenetic analyses of gene trees for transcripts annotated MRPP1-3 confirm they being homologous to MRPP1-3 genes derived from other organisms (Figs. S3-5). Lopez Sanchez *et al.* (2011) recently showed that MRPP1 and MRPP3 are sufficient for processing the 5' end of mitochondrial tRNAs in humans.<sup>49</sup>

The same authors and others identified the zinc phosphodiesterase ELAC protein 2 as RNase Z responsible for processing the 3' end of mitochondrial tRNAs.<sup>49,50</sup> We recovered two transcripts that encode for proteins similar to ELAC2: the shorter transcript (ELAC2a) displayed more similarity (in the form of higher matching base and alignment length BLAST results in the Uniprot database) with homologs from other taxa but was incomplete, missing the 5' terminal; the longer transcript (ELAC2b) lacked a mitochondrial targeting signal peptide (Table 2). Finally, we identified one CCAsel homolog that lacked an acceptable mitochondrial targeting signal peptide (Table 2).

**Table 3.** List of aminoacyl-tRNA synthetases (aaRS) found in the RNA-seq data of *Alatina alata*.

name	CDS size <sup>¶</sup>	mt-signal	mt-sign size <sup>§§</sup>	Mito ProtII	Target P	Predotar	Mito Fates
A	2985	+	29/29	<b>0.97</b>	<b>0.89</b>	<b>0.81</b>	<b>0.99</b>
C	2343	+	16/100	<b>0.62</b>	<b>0.56</b>	0.47	0.22
D	1803	+	72/30	<b>0.99</b>	<b>0.81</b>	0.43	<b>0.69</b>
E/P	4371	–	NA	0.04	0.11	0.01	0.00
mt-F	1356	+	37/37	<b>0.99</b>	<b>0.76</b>	0.02	0.46
nc-Fa	1512	–	26/–	0.13	0.08	0.00	0.01
nc-Fb	1782	–	NA	0.25	0.06	0.00	0.00
G	2283	+	44/35	<b>0.99</b>	<b>0.59</b>	<b>0.72</b>	<b>0.60</b>
H	>1578	?	NA	NA	NA	NA	NA
I	3543	–	NA	0.01	0.06	0.00	0.02
K	1893	+	35/42	<b>0.82</b>	<b>0.61</b>	<b>0.68</b>	<b>0.61</b>
L	3294	+	31/22	<b>0.99</b>	0.47	<b>0.52</b>	<b>0.72</b>
M	2970	+	28/12	<b>0.98</b>	<b>0.94</b>	<b>0.72</b>	<b>0.92</b>
N	1896	+	39/–	<b>0.97</b>	0.50	0.10	<b>0.83</b>
Q	2427	+	26/26	<b>0.97</b>	<b>0.84</b>	<b>0.51</b>	<b>0.78</b>
R	2175	+	55/56	<b>1.00</b>	<b>0.90</b>	<b>0.90</b>	<b>1.00</b>
S	1617	+	14/27	<b>0.98</b>	<b>0.93</b>	<b>0.66</b>	<b>0.62</b>
T	2172	+	50/50	<b>0.99</b>	<b>0.72</b>	0.29	<b>0.84</b>
V	3288	–	NA/–	0.14	0.50	0.38	<b>0.79</b>
W	1413	+	27/47	<b>0.99</b>	<b>0.76</b>	0.41	<b>0.89</b>
Y	1797	+	57/58	<b>0.99</b>	<b>0.89</b>	<b>0.86</b>	<b>0.58</b>

Aminoacyl-tRNA synthetases are listed according to their one letter name. Presence of mt-signal is indicated when at least two prediction software display probabilities >0.50 (in bold). mt-F: monomeric mitochondrial aaRS for phenylalanine; nc-Fa and nc-Fb:  $\alpha$  and  $\beta$  subunits of the dimeric cytoplasmic aaRS for phenylalanine. CDS size<sup>¶</sup>: size of the coding regions found in the transcriptome assembly; mt-sign size<sup>§§</sup>: number of amino acids composing the mt-targeting signal peptide as predicted by MitoProt II v1.01 and TargetP v1.1. Probabilities of mitochondrial targeting were obtained using MitoProt II v1.01, TargetP v1.1, Predotar v.1.03 and MitoFates.

The presence of mitochondrial targeting signals on the transcripts of several mitochondrial tRNA-processing genes combined with the high expression of mt-tRNA<sup>Met</sup> (Table S3) suggests that *trnM* requires a functional tRNA structure in the mitochondrion of this species (Fig. 3B). Given the very distinct bacterial-like translation initiation system in the mitochondrion compared to the eukaryote-like one for the cytosol, the loss of *trnM* is thought to be nigh impossible.<sup>51</sup> Yet, the loss of mt-MetRS requires nc-MetRS, whose transcript harbors an mt-targeting signal peptide (Table 3) to recognize and aminoacetylate mt-tRNA<sup>Met</sup> in the mitochondrion. Alternatively, *trnM* could play a role as initiation of translation in the mitochondrion.

Both Octocorallia and Cubozoa have lost both the mitochondrial tRNA and the mitochondrial aminoacyl-tRNA synthetase for tryptophan (mt-tRNA<sup>Trp</sup> and mt-TrpRS, respectively).<sup>29</sup> In the absence of both mt-tRNA<sup>Trp</sup> and mt-TrpRS, tryptophan residues are processed by a combination of cytoplasmic nc-tRNA<sup>Trp</sup> and nc-TrpRS. In the ctenophore *Pleurobrachia bachei*, the loss of mt-tRNA<sup>Trp</sup> has resulted in the reassignment of the TGA codon to serine.<sup>29,52,53</sup> In cubozoan mtDNA, tryptophan residues are encoded by two codons TGG and TGA at a relatively similar proportion (Table S3). In octocorals, TGG is the main codon used for tryptophan, the exceptions seemingly resulting either from sequencing errors or present in the extra protein coding gene homolog to bacterial



**Table 4.** List of mitochondrial ribosomal proteins (MRPs) found in the RNA-seq data of *Alatina alata*.

MRPs	CDS size <sup>†</sup>	mt-signal	mt-sign size <sup>‡</sup>	MitoProtII	TargetP	Predotar	MitoFates
L2	966	+	22/69	<b>0.8928</b>	<b>0.894</b>	0,41	<b>0.883</b>
L3	1131	+	46/39	<b>0.9912</b>	<b>0.831</b>	0,42	<b>0.540</b>
L4	909	+	47/26	<b>0.9969</b>	<b>0.884</b>	<b>0,66</b>	<b>0.989</b>
L9	681	+	34/9	<b>0.7681</b>	<b>0.560</b>	0,50	<b>0.655</b>
L10	651	+	51/109	<b>0.9916</b>	<b>0.647</b>	0,50	<b>0.800</b>
L11	459	–	–/–	0.3073	0.216	0,44	0.039
L12	597	+	46/46	<b>0.9672</b>	<b>0.729</b>	0,22	<b>0.620</b>
L13	510	–	–/103	0.3097	<b>0.758</b>	0,30	0.068
L14	576	+	111/27	<b>0.9548</b>	<b>0.759</b>	0,47	<b>0.615</b>
L15	795	–	28/–	<b>0.6694</b>	0.400	0,15	0.488
L17	486	+	–/13	<b>0.7122</b>	<b>0.718</b>	0,21	0.450
L18	690	+	45/45	<b>0.6141</b>	<b>0.842</b>	<b>0,55</b>	<b>0.698</b>
L19	984	–	–/–	<b>0.6420</b>	0.190	0,09	0.175
L20	381	+	71/38	<b>0.9999</b>	<b>0.876</b>	<b>0,64</b>	<b>0.878</b>
L21	705	+	45/71	<b>0.9902</b>	<b>0.890</b>	<b>0,55</b>	<b>0.818</b>
L22	612	+	30/30	<b>0.9904</b>	<b>0.924</b>	<b>0,94</b>	<b>0.996</b>
L23	471	+	–/5	<b>0.8558</b>	<b>0.812</b>	0,12	0.500
L24	738	+	21/119	<b>0.9780</b>	<b>0.641</b>	<b>0,68</b>	0.365
L28	519	+	33/15	<b>0.9809</b>	<b>0.616</b>	<b>0,77</b>	<b>0.906</b>
L29/L47	597	+	28/64	<b>0.9374</b>	<b>0.872</b>	<b>0,52</b>	<b>0.694</b>
L30	429	–	–/–	<b>0.6505</b>	0.272	0,21	0.013
L32	318	+	13/50	<b>0.5893</b>	0.466	0,26	<b>0.660</b>
L33	378	–	–/15	0.4681	0.079	0,00	0.026
L34	264	+	23/41	<b>0.9761</b>	<b>0.787</b>	<b>0,88</b>	<b>0.950</b>
L35	480	+	47/109	<b>0.9755</b>	<b>0.928</b>	<b>0,53</b>	<b>0.796</b>
L36	216	+	50/39	<b>0.9912</b>	<b>0.717</b>	0,35	<b>0.794</b>
L37a	1053	+	–/35	<b>0.9135</b>	<b>0.675</b>	0,09	0.148
L37b	1143	+	11/36	<b>0.9610</b>	<b>0.794</b>	<b>0,52</b>	0.380
L38	975	+	–/29	<b>0.5728</b>	<b>0.753</b>	0,01	<b>0.915</b>
L40	540	–	17/–	<b>0.9278</b>	0.151	0,48	0.006
L41	303	+	32/42	<b>0.8645</b>	<b>0.595</b>	<b>0,63</b>	<b>0.871</b>
L43	423	–	–/–	<b>0.9212</b>	0.262	0,22	0.377
L44	1002	+	19/12	<b>0.9418</b>	<b>0.706</b>	<b>0,67</b>	<b>0.751</b>
L45	801	+	35/26	<b>0.9767</b>	<b>0.858</b>	0,47	0.062
L46	909	+	11/13	<b>0.7244</b>	<b>0.624</b>	0,32	<b>0.840</b>
L48	600	+	13/45	<b>0.9404</b>	<b>0.655</b>	<b>0,52</b>	<b>0.775</b>
L49	468	+	25/44	<b>0.9936</b>	<b>0.907</b>	0,46	0.196
L52	345	–	35/–	<b>0.8568</b>	0.391	0,47	0.354
L54	357	+	19/19	<b>0.9452</b>	<b>0.631</b>	0,26	<b>0.829</b>
L55	420	+	–/6	<b>0.8509</b>	<b>0.777</b>	0,44	0.482
S2	810	+	21/36	<b>0.9895</b>	<b>0.944</b>	<b>0,77</b>	<b>0.945</b>
S3/S24	495	+	46/28	<b>0.9719</b>	<b>0.815</b>	<b>0,73</b>	<b>0.750</b>
S5	1281	+	14/–	<b>0.8815</b>	0.448	0,23	<b>0.828</b>
S6	363	–	–/–	0.2581	0.094	0,05	0.132
S7	711	+	15/–	<b>0.8070</b>	<b>0.748</b>	<b>0,83</b>	0.427
S9	1170	+	38/63	<b>0.9000</b>	<b>0.643</b>	0,01	<b>0.914</b>
S10	747	+	37/48	<b>0.9680</b>	<b>0.839</b>	0,03	<b>0.976</b>
S11	648	+	61/61	<b>0.9905</b>	0.490	0,39	<b>0.613</b>
S12	606	+	153/83	<b>0.9986</b>	<b>0.770</b>	<b>0,81</b>	<b>0.799</b>
S14	486	+	14/53	<b>0.6913</b>	<b>0.816</b>	0,41	<b>0.751</b>
S15	657	+	14/101	<b>0.6887</b>	<b>0.664</b>	0,45	<b>0.719</b>
S16	375	+	37/10	<b>0.9752</b>	<b>0.846</b>	<b>0,81</b>	<b>0.553</b>
S17	279	–	–/–	<b>0.9399</b>	0.144	0,10	0.028
S18A	393	+	26/17	<b>0.9935</b>	<b>0.839</b>	<b>0,59</b>	<b>0.985</b>
S18B	546	+	38/31	<b>0.9879</b>	<b>0.959</b>	0,40	<b>0.983</b>
S18C	501	+	38/55	<b>0.8529</b>	<b>0.713</b>	0,22	0.433
S21	270	+	40/28	<b>0.6619</b>	<b>0.622</b>	0,06	0.135
S22	1068	+	45/29	<b>0.9663</b>	<b>0.799</b>	<b>0,57</b>	0.248
S23	567	+	15/6	<b>0.6029</b>	0.458	0,11	<b>0.668</b>
S25a	492	+	–/52	<b>0.5498</b>	0.425	0,07	<b>0.526</b>
S25b	483	+	–/–	<b>0.7713</b>	0.468	<b>0,64</b>	0.111
S26	771	+	27/19	<b>0.9902</b>	<b>0.905</b>	<b>0,58</b>	<b>0.810</b>
S28	498	–	33/–	<b>0.8869</b>	0.353	0,26	0.387
S29	1215	+	37/38	<b>0.9914</b>	<b>0.824</b>	0,22	0.480
S31	2307	+	30/48	<b>0.8050</b>	<b>0.781</b>	<b>0,86</b>	0.500
S33	312	+	31/38	<b>0.8297</b>	<b>0.589</b>	0,02	0.477
S35	1047	–	39/–	<b>0.6656</b>	0.327	0,00	0.159
S36	376	+	21/20	<b>0.9792</b>	<b>0.802</b>	<b>0,57</b>	<b>0.650</b>
S39	2043	+	47/40	<b>0.8781</b>	<b>0.832</b>	0,49	<b>0.944</b>
MRP63	354	+	27/–	<b>0.9871</b>	0.444	<b>0,89</b>	0.305

Presence of mt-signal is indicated when at least two prediction software display probabilities >0.50 (in bold). CDS size<sup>†</sup>: size of the coding regions found in the transcriptome assembly; mt-sign size<sup>‡</sup>: number of amino acids composing the mt-targeting signal peptide as predicted by MitoProt II v1.01 and TargetP v1.1. Probabilities of mitochondrial targeting were obtained using MitoProt II v1.01, TargetP v1.1, Predotar v1.03 and MitoFates.

*mutS*<sup>54</sup> (Table S3). In the cytoplasm, nc-tRNA<sup>Trp</sup> only recognizes TGG, while TGA is a stop codon. In the absence of mRNA editing in *Alatina* (proportion of TGA was similar in RNA and DNA-seq data; Table S3), we suggest that the anticodon of nc-tRNA<sup>Trp</sup>(CCA) is edited in the mitochondrion to recognize and aminoacylate TGA codons (anticodon UCA) through a C → U editing mechanism (Fig. 3C). Anticodon C to U editing has been reported in the mitochondrion of *Trypanosoma* allowing the recognition of TGA as tryptophan in the absence of an mt-tRNA<sup>Trp</sup>(UCA) gene,<sup>55</sup> and in marsupials where some of the mt-tRNA<sup>Asp</sup>(GCC) are edited to mt-tRNA<sup>Asp</sup>(GUC).<sup>56</sup> Alternatively, addition factors (RNA and/or proteins) would be required to guide TGA recognition by a native nc-tRNA<sup>Trp</sup>. Future studies focusing on short RNAs should allow testing our hypotheses.

### The mitochondrial ribosomal proteins complex

Ribosomes, essential pieces of the translation machinery, are complex molecules made of two subunits (large and small), each composed of RNA (of different origin for the cytoplasm and the mitochondrion) and many proteins. In animals, all mitochondrial proteins are encoded by the nucleus, while two RNA subunits (12S and 16S for small and large RNA subunits, respectively) are encoded by the mitochondrial DNA (*rns* and *rnl*). We used the dataset of a recent study on Holozoa<sup>57</sup> to investigate ribosomal protein coding genes in the transcriptome of *A. alata*. We recovered the same cnidarian set of large (40 gene transcripts) and small (29 gene transcripts) subunits of the mitochondrial ribosome, with a few exceptions, plus the unassigned MRP63 (Fig. 3D; Table 4).

From the large ribosomal subunit, L1, L16, L27, L50, L53 and L56 were not recovered. From the small ribosomal subunit, only S30 was not recovered. We also found that the subunits L11, L13, L15, L19, L30, L33, L40, L43 and L52 from the large subunit, and S6, S17, S28 and S35 from the small subunit had low probability ( $\leq 0.5$  for at least three out of the four prediction software used) for possessing mitochondrial-targeting signal peptides (Table 4). Sheel *et al.* (2014) also found L50 and L56 to be absent from the hydrozoan *Hydra magnapapillata*.<sup>57</sup> Assuming the lack of these genes reflects their loss in the cubozoan *A. alata*, the loss of L56 could be a synapomorphy of Medusozoa (*Hydra* and *Alatina*), while the loss of L1, L16, L27 and S30 appear to be autapomorphies to *Alatina*.

Interestingly, we found two versions of subunits L37 and S25, suggesting gene duplication, though an assembly artifact is conceivable. Both candidates for subunit L37 had recognizable mitochondrial-targeting signal peptides but the longer transcript (L37b) displayed higher probability of export to the mitochondrion, while the shorter transcript (L37a) showed more similarity (in the form of higher matching base and alignment length BLAST results in the Uniprot database) to homologs from other taxa (Table 4). Similarly, while we identified a mitochondrial-targeting signal peptide on the two candidate transcripts for subunit S25, the longer transcript (S25b) showed more similarity to homologs from other taxa (Table 4). We also found the transcript for subunit S31 to be far larger in size than homologs from other metazoans, with an addition of ~840 nucleotides (~280 amino acids) on the 5'-end of the coding

DNA sequence (CDS). The function of the extra amino acids in S31 was not determined.

Lastly, we identified another mitochondrial ribosomal protein subunit (S39) with similarities to pentatricopeptide repeat-containing protein 3 previously not found by Sheel and Hausdorf (2014).<sup>57</sup> Using the BLAST search tool on the Uniprot database, we identified homologs from *Nematostella vectensis* (Cnidaria: Anthozoa; predicted protein v1g212337), *Hydra vulgaris* (Cnidaria: Hydrozoa; pentatricopeptide repeat-containing protein 3, mitochondrial T2M872) and *Amphimedon queenslandica* (Porifera: Demospongiae; uncharacterized protein LOC100633076) with mitochondrial targeting signal peptides (data not shown).

### Conclusion

Cubozoans have one of the most unusual mtDNA organizations in Metazoa, where genes are distributed on eight linear chromosomes with long terminal inverted repeats. We explored the recently published RNA-seq data from the box jellyfish *Alatina alata* to better understand the mechanisms involved in the expression of linear mitochondrial chromosomes. We propose a pattern of mitochondrial gene expression where each mitochondrion is transcribed into a single transcript starting at a well conserved origin of transcription (OT; 30 bp) and ending with the polyadenylation site about 1/4<sup>th</sup> of the way into the inverted repeat (IR) downstream of the coding regions of the mitochondrion. We captured the two extra protein-coding genes *mt-polB* and *orf314* in the RNA-seq data with coverage levels similar to other mitochondrial protein-coding genes, providing further evidence for the functionality of these genes. We also found that the evolution of mitochondrial translational machinery in *A. alata* displays similarities to what has been documented in other non-bilaterians.<sup>29</sup> Finally, we suggest that the mt-tRNA<sup>Met</sup> is likely a functional tRNA that could be recognized and aminoacylated by the nc-MetRS, and that an imported nc-tRNA<sup>Trp</sup> is edited in order to recognize TGA codons as tryptophan. Future studies exploring the repertoire of small RNAs and the proteome of the mitochondrion will be able to test these hypotheses.

### Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

### Acknowledgments

Some of this work was performed using resources of the Laboratories of Analytical Biology at the Smithsonian National Museum of Natural History. EK and BB were supported by a Smithsonian Institution Peter Buck postdoctoral fellowship. We also want to thank two anonymous reviewers for providing constructive criticisms to earlier versions of this manuscript.

### References

- Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, Podowski RM, Naslund A, Eriksson A-SS, Winkler HH, Kurland CG, *et al.* The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 1998; 396:133-40; PMID:9823893; <http://dx.doi.org/10.1038/24094>

2. Pittis AA, Gabaldón T. Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature* 2016; 531:101-4; PMID:26840490; <http://dx.doi.org/10.1038/nature16941>
3. Andersson SG, Karlberg O, Canbäck B, Kurland CG. On the origin of mitochondria: a genomics perspective. *Philos Trans R Soc Lond B Biol Sci* 2003; 358:165-77; PMID:12594925; <http://dx.doi.org/10.1098/rstb.2002.1193>
4. Holt IJ, Reyes A. Human mitochondrial DNA Replication. *Cold Spring Harb Perspect Biol* 2012; 4:a012971; <http://dx.doi.org/10.1101/cshperspect.a012971>
5. Ojala D, Montoya J, Attardi G. tRNA punctuation model of RNA processing in human mitochondria. *Nature* 1981; 290:470-4; PMID:7219536; <http://dx.doi.org/10.1038/290470a0>
6. Brugler MR. Molecular evolution in black corals (Cnidaria Anthozoa Hexacorallia) implications for antipatharian systematics. 2011.
7. Shao Z, Graf S, Chaga OY, Lavrov DV. Mitochondrial genome of the moon jelly *Aurelia aurita* (Cnidaria, Scyphozoa): A linear DNA molecule encoding a putative DNA-dependent DNA polymerase. *Gene* 2006; 381:92-101; PMID:16945488; <http://dx.doi.org/10.1016/j.gene.2006.06.021>
8. Voigt O, Erpenbeck D, Wörheide G. A fragmented metazoan organellar genome: the two mitochondrial chromosomes of *Hydra magnipillata*. *BMC Genomics* 2008; 9:350; PMID:18655725; <http://dx.doi.org/10.1186/1471-2164-9-350>
9. Kayal E, Lavrov DV. The mitochondrial genome of *Hydra oligactis* (Cnidaria, Hydrozoa) sheds new light on animal mtDNA evolution and cnidarian phylogeny. *Gene* 2008; 410:177-86; PMID:18222615; <http://dx.doi.org/10.1016/j.gene.2007.12.002>
10. Park E, Hwang D-S, Lee J-S, Song J-I, Seo T-K, Won Y-J. Estimation of divergence times in cnidarian evolution based on mitochondrial protein-coding genes and the fossil record. *Mol Phylogenet Evol* 2012; 62:329-45; PMID:22040765; <http://dx.doi.org/10.1016/j.ympev.2011.10.008>
11. Zou H, Zhang J, Li W, Wu S, Wang G. Mitochondrial Genome of the Freshwater Jellyfish *Craspedacusta sowerbyi* and Phylogenetics of Medusozoa. *PLoS One* 2012; 7:e51465; PMID:23240028; <http://dx.doi.org/10.1371/journal.pone.0051465>
12. Kayal E, Bentlage B, Collins AG, Kayal M, Pirro S, Lavrov DV. Evolution of linear mitochondrial genomes in medusozoan cnidarians. *Genome Biol Evol* 2012; 4:1-12; PMID:22113796; <http://dx.doi.org/10.1093/gbe/evr123>
13. Smith DR, Kayal E, Yanagihara AA, Collins AG, Pirro S, Keeling PJ. First complete mitochondrial genome sequence from a box jellyfish reveals a highly fragmented, linear architecture and insights into telomere evolution. *Genome Biol Evol* 2012; 4:52-8; PMID:22117085; <http://dx.doi.org/10.1093/gbe/evr127>
14. Luz BLP, Capel KCC, Stampar SN, Kitahara MV. Description of the mitochondrial genome of the tree coral *Dendrophyllia arbuscula* (Anthozoa, Scleractinia). *Mitochondrial DNA* 2015; 27(4):1-2; PMID:26119126; <http://dx.doi.org/10.3109/19401736.2015.1060435>
15. Kayal E, Bentlage B, Cartwright P, Yanagihara AA, Lindsay DJ, Hopcroft RR, Collins AG. Phylogenetic analysis of higher-level relationships within Hydrozoa (Cnidaria: Hydrozoa) using mitochondrial genome data and insight into their mitochondrial transcription. *PeerJ* 2015; 3:e1403; PMID:26618080; <http://dx.doi.org/10.7717/peerj.1403>
16. Arrigoni R, Vacherie B, Benzoni F, Barbe V. The complete mitochondrial genome of *Acanthastrea maxima* (Cnidaria, Scleractinia, Lobophylliidae). *Mitochondrial DNA* 2016; 27(2):927-8; PMID:24938099; <http://dx.doi.org/10.3109/19401736.2014.926489>
17. Wu J-S, Ju Y-M, Hsiao S-T, Hsu C-H. Complete mitochondrial genome of *Junceella fragilis* (Gorgonacea, Ellisellidae). *Mitochondrial DNA* 2016; 27(2):1229-30; PMID:25090396; <http://dx.doi.org/10.3109/19401736.2014.945531>
18. Figueroa DF, Baco AR. Complete mitochondrial genomes elucidate phylogenetic relationships of the deep-sea octocoral families Coralliidae and Paragorgiidae. *Deep Res Part II Top Stud Oceanogr* 2014; 99:83-91; <http://dx.doi.org/10.1016/j.dsr2.2013.06.001>
19. Foxx J, Brugler MR, Siddall ME, Rodríguez E. Multiplexed pyrosequencing of nine sea anemone (Cnidaria: Anthozoa: Hexacorallia: Actiniaria) mitochondrial genomes. *Mitochondrial DNA* 2015; 27(4):1-7; PMID:26104159; <http://dx.doi.org/10.3109/19401736.2015.1053114>
20. Ender A, Schierwater B. Placozoa are not derived cnidarians: evidence from molecular morphology. *Mol Biol Evol* 2003; 20:130-4; PMID:12519915; <http://dx.doi.org/10.1093/molbev/msg018>
21. Mouhamadou B, Barroso G, Labarère J. Molecular evolution of a mitochondrial polB gene, encoding a family B DNA polymerase, towards the elimination from *Agrocybe* mitochondrial genomes. *Mol Genet Genomics* 2004; 272:257-63; PMID:15365817; <http://dx.doi.org/10.1007/s00438-004-1050-4>
22. Oldenburg DJ, Bendich AJ. Mitochondrial DNA from the liverwort *Marchantia polymorpha*: circularly permuted linear molecules, head-to-tail concatemers, and a 5' protein. *J Mol Biol* 2001; 310:549-62; PMID:11439023; <http://dx.doi.org/10.1006/jmbi.2001.4783>
23. Tomáška L, Nosek J, Fukuhara H. Identification of a putative mitochondrial telomere-binding protein of the yeast *Candida parapsilosis*. *J Biol Chem* 1997; 272:3049-56; PMID:9006955; <http://dx.doi.org/10.1074/jbc.272.5.3049>
24. Fricova D, Valach M, Farkas Z, Pfeiffer I, Kucsera J, Tomáška L, Nosek J. The mitochondrial genome of the pathogenic yeast *Candida subhashii*: GC-rich linear DNA with a protein covalently attached to the 5' termini. *Microbiology* 2010; 156:2153-63; PMID:20395267; <http://dx.doi.org/10.1099/mic.0.038646-0>
25. Zapata F, Goetz FE, Smith SA, Howison M, Siebert S, Church SH, Sanders SM, Ames CL, McFadden CS, France SC, et al. Phylogenomic analyses support traditional relationships within Cnidaria. *PLoS One* 2015; 10:e0139068; PMID:26465609; <http://dx.doi.org/10.1371/journal.pone.0139068>
26. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012; 9:357-9; PMID:22388286; <http://dx.doi.org/10.1038/nmeth.1923>
27. Kears M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 2012; 28:1647-9; PMID:22543367; <http://dx.doi.org/10.1093/bioinformatics/bts199>
28. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; 30:2114-20; PMID:24695404; <http://dx.doi.org/10.1093/bioinformatics/btu170>
29. Pett W, Lavrov DV. Cytonuclear interactions in the evolution of animal mitochondrial tRNA metabolism. *Genome Biol Evol* 2015; 7:2089-2101; PMID:26116918; <http://dx.doi.org/10.1093/gbe/evv124>
30. Chevreux B, Wetter T, Suhai S. Genome sequence assembly using trace signals and additional sequence information. *Comput Sci Biol Proc Ger Conf Bioinforma* 1999; 99:45-56
31. Hahn C, Bachmann L, Chevreux B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res* 2013; 41:e129; PMID:23661685; <http://dx.doi.org/10.1093/nar/gkt371>
32. Claros MG, Vincens P. Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur J Biochem* 1996; 241:779-86; PMID:8944766; <http://dx.doi.org/10.1111/j.1432-1033.1996.00779.x>
33. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* 2007; 2:953-71; PMID:17446895; <http://dx.doi.org/10.1038/nprot.2007.131>
34. Small I, Peeters N, Legeai F, Lurin C. Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* 2004; 4:1581-90; PMID:15174128; <http://dx.doi.org/10.1002/pmic.200300776>
35. Fukasawa Y, Tsuji J, Fu S-C, Tomii K, Horton P, Imai K. MitoFates: Improved Prediction of Mitochondrial Targeting Sequences and Their Cleavage Sites. *Mol Cell Proteomics* 2015; 14:1113-26; PMID:25670805; <http://dx.doi.org/10.1074/mcp.M114.043083>
36. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013; 30:772-80; PMID:23329690; <http://dx.doi.org/10.1093/molbev/mst010>
37. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000; 17:540-52;

- PMID:10742046; <http://dx.doi.org/10.1093/oxfordjournals.molbev.a026334>
38. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014; 30:1312-3; PMID:24451623; <http://dx.doi.org/10.1093/bioinformatics/btu033>
  39. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 2015; 44:D279-85; PMID:26673716; <http://dx.doi.org/10.1093/nar/gkv1344>
  40. Eddy SR. Profile hidden Markov models. *Bioinformatics* 1998; 14:755-63; PMID:9918945; <http://dx.doi.org/10.1093/bioinformatics/14.9.755>
  41. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003; 31:3406-15; PMID:12824337; <http://dx.doi.org/10.1093/nar/gkg595>
  42. Fernández-Silva P, Enriquez JA, Montoya J. Replication and transcription of mammalian mitochondrial DNA. *Exp Physiol* 2003; 88:41-56; PMID:12525854; <http://dx.doi.org/10.1113/eph8802514>
  43. Haen KM, Pett W, Lavrov DV. Parallel loss of nuclear-encoded mitochondrial aminoacyl-tRNA synthetases and mtDNA-encoded tRNAs in Cnidaria. *Mol Biol Evol* 2010; 27:2216-9; PMID:20439315; <http://dx.doi.org/10.1093/molbev/msq112>
  44. Brown A, Amunts A, Bai X, Sugimoto Y, Edwards PC, Murshudov G, Scheres SH, Ramakrishnan V. Structure of the large ribosomal subunit from human mitochondria. *Science* 2014; 346:718-22; PMID:25278503; <http://dx.doi.org/10.1126/science.1258026>
  45. Greber BJ, Boehringer D, Leitner A, Bieri P, Voigts-Hoffmann F, Erzberger JP, Leibundgut M, Aebersold R, Ban N. Architecture of the large subunit of the mammalian mitochondrial ribosome. *Nature* 2014; 505:515-9; PMID:24362565; <http://dx.doi.org/10.1038/nature12890>
  46. Wang X, Lavrov DV. Seventeen new complete mtDNA sequences reveal extensive mitochondrial genome evolution within the demospongiae. *PLoS One* 2008; 3:e2723; PMID:18628961; <http://dx.doi.org/10.1371/journal.pone.0002723>
  47. Holzmann J, Frank P, Löffler E, Bennett KL, Gerner C, Rossmannith W. RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. *Cell* 2008; 135:462-74; PMID:18984158; <http://dx.doi.org/10.1016/j.cell.2008.09.013>
  48. Rossmannith W, Holzmann J. Processing mitochondrial (t)RNAs: New enzyme, old job. *Cell Cycle* 2009; 8:1650-3; PMID:19411849; <http://dx.doi.org/10.4161/cc.8.11.8502>
  49. Lopez Sanchez MI, Mercer TR, Davies SM, Shearwood A-MJ, Nygård KK, Richman TR, Mattick JS, Rackham O, Filipovska A. RNA processing in human mitochondria. *Cell Cycle* 2011; 10:2904-16; PMID:21857155; <http://dx.doi.org/10.4161/cc.10.17.17060>
  50. Brzezniak LK, Bijata M, Szczesny RJ, Stepien PP. Involvement of human ELAC2 gene product in 3' end processing of mitochondrial tRNAs. *RNA Biol* 2011; 8:616-26; PMID:21593607; <http://dx.doi.org/10.4161/rna.8.4.15393>
  51. Schneider A. Mitochondrial tRNA import and its consequences for mitochondrial translation. *Annu Rev Biochem* 2010; 80:1-21; PMID:21417719; <http://dx.doi.org/10.1146/annurev-biochem-060109-092838>
  52. Kohn AB, Citarella MR, Kocot KM, Bobkova YV, Halanych KM, Moroz LL. Rapid evolution of the compact and unusual mitochondrial genome in the ctenophore, *Pleurobrachia bachei*. *Mol Phylogenet Evol* 2011; 63:203-7; PMID:22201557; <http://dx.doi.org/10.1016/j.ympev.2011.12.009>
  53. Pett W, Ryan JF, Pang K, Mullikin JC, Martindale MQ, Baxeavanis AD, Lavrov DV. Extreme mitochondrial evolution in the ctenophore *Mnemiopsis leidyi*: Insight from mtDNA and the nuclear genome. *Mitochondrial DNA* 2011; 22(4):1-13; PMID:21985407; <http://dx.doi.org/10.3109/19401736.2011.624611>
  54. Brugler MR, France SC. The mitochondrial genome of a deep-sea bamboo coral (Cnidaria, Anthozoa, Octocorallia, Isididae): genome structure and putative origins of replication are not conserved among octocorals. *J Mol Evol* 2008; 67:125-36; PMID:18506498; <http://dx.doi.org/10.1007/s00239-008-9116-2>
  55. Alfonzo JD, Blanc V, Estévez AM, Rubio MA, Simpson L. C to U editing of the anticodon of imported mitochondrial tRNA(Trp) allows decoding of the UGA stop codon in *Leishmania tarentolae*. *EMBO J* 1999; 18:7056-62; PMID:10601027; <http://dx.doi.org/10.1093/emboj/18.24.7056>
  56. Janke A, Pääbo S. Editing of a tRNA anticodon in marsupial mitochondria changes its codon recognition. *Nucleic Acids Res* 1993; 21:1523-5; PMID:8479901; <http://dx.doi.org/10.1093/nar/21.7.1523>
  57. Scheel BM, Hausdorf B. Dynamic evolution of mitochondrial ribosomal proteins in Holozoa. *Mol Phylogenet Evol* 2014; 76:67-74; PMID:24631858; <http://dx.doi.org/10.1016/j.ympev.2014.03.005>